



## De novo production of benzyl glucosinolate in *Escherichia coli*

Petersen, Annette; Crocoll, Christoph; Halkier, Barbara Ann

*Published in:*  
Metabolic Engineering

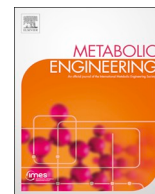
*DOI:*  
[10.1016/j.ymben.2019.02.004](https://doi.org/10.1016/j.ymben.2019.02.004)

*Publication date:*  
2019

*Document version*  
Publisher's PDF, also known as Version of record

*Document license:*  
[CC BY-NC-ND](#)

*Citation for published version (APA):*  
Petersen, A., Crocoll, C., & Halkier, B. A. (2019). De novo production of benzyl glucosinolate in *Escherichia coli*. *Metabolic Engineering*, 54, 24-34. <https://doi.org/10.1016/j.ymben.2019.02.004>



# De novo production of benzyl glucosinolate in *Escherichia coli*

Annette Petersen, Christoph Crocoll, Barbara Ann Halkier\*

DynaMo Center, Copenhagen Plant Science Centre, Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark

## ARTICLE INFO

### Keywords:

Glucosinolates  
*Escherichia coli*  
Metabolic engineering  
Targeted proteomics

## ABSTRACT

Microbial production of plant specialised metabolites is challenging as the biosynthetic pathways are often complex and can contain enzymes, which function is not supported in traditional production hosts. Glucosinolates are specialised metabolites of strong commercial interest due to their health-promoting effects. In this work, we engineered the production of benzyl glucosinolate in *Escherichia coli*. We systematically optimised the production levels by first screening different expression strains and by modification of growth conditions and media compositions. This resulted in production from undetectable to approximately 4.1  $\mu\text{M}$  benzyl glucosinolate, but also approximately 3.7  $\mu\text{M}$  of desulfo-benzyl glucosinolate, the final intermediate of this pathway. Additional optimisation of pathway flux through entry point cytochrome P450 enzymes and PAPS-dependent sulfotransferase increased the production additionally 5-fold to 20.3  $\mu\text{M}$  (equivalent to 8.3 mg/L) benzyl glucosinolate.

## 1. Introduction

A continuously increasing number of plant specialised metabolites are considered high value compounds due to e.g. pharmaceutical, nutraceutical or cosmetic properties. Production through natural sources is often not sustainable because of low abundance of compound and/or source as well as expensive purification processes (Ikram et al., 2015). The demand has inspired researchers to apply metabolic engineering and synthetic biology approaches to microorganisms. The efforts have led to innumerable reports on heterologous production of plant specialised metabolites in expression hosts such as *Escherichia coli* and *Saccharomyces cerevisiae* (Chae et al., 2017; Cho et al., 2015; Huang et al., 2008; Pontrelli et al., 2018; Siddiqui et al., 2012). Only few compounds have reached commercially relevant production levels and production is most commonly reported in the  $\mu\text{g/L}$  to low mg/L range (Chubukov et al., 2016; Pontrelli et al., 2018). In pursuit of still higher titres, researchers continue to develop technologies in the fields of omics, high-throughput analytics and computational engineering. Today, strategies to increase yields often include multiple rounds of optimisation and large mutant library screens - an approach also known as systems metabolic engineering (Chae et al., 2017; Cho et al., 2015; Yang et al., 2017). An emerging approach to identify bottlenecks and optimise flux through a pathway is mass spectrometry-based targeted proteomics by selected reaction monitoring (SRM). By using isotopically labelled synthetic

peptides as internal references in the samples, it is possible to determine not only the presence but also the relative or absolute quantities of a given protein of interest e.g. over time or at different conditions (Batth et al., 2012). This method has been used successfully for increasing production titres of several products in *E. coli* (Alonso-Gutierrez et al., 2015; Batth et al., 2012; Brunk et al., 2016; Dahl et al., 2013; George et al., 2015, 2014; Juminaga et al., 2012; Redding-Johanson et al., 2011; Singh et al., 2012; Tan et al., 2016).

One group of plant specialised metabolites with commercial interest is glucosinolates (GLSs). They are sulfur-rich compounds characteristic of brassicaceous plants, including the model plant *Arabidopsis thaliana* and several agriculturally important crops (Fahey et al., 2001). In plants, GLSs constitute a line of defence, which becomes activated upon attack from pathogens or herbivores (Jeschke and Burow, 2018). For humans, the intake of brassicaceous vegetables has been associated with reduced risks of cardiovascular diseases and several types of cancer (reviewed in (Traka, 2016)), and with better insulin resistance for type 2 diabetes (Bahadoran et al., 2012). These health beneficial effects have made GLSs interesting as dietary supplements and primed a desire to engineer microbial production as rich source of GLSs.

GLSs are derived from amino acids and synthesised through a core structure pathway consisting of seven enzymatic steps (Fig. 1). Some GLSs, e.g. the aliphatic methionine-derived GLSs, require chain

**Abbreviations:** dsBGLS, desulfo-benzyl glucosinolate; BGLS, benzyl glucosinolate; I3M, indol-3-ylmethyl glucosinolate; 4MSB, 4-methylsulfinylbutyl glucosinolate; APS, adenosine 5'-phosphosulfate; PAPS, 3'-phosphoadenosine-5'-phosphosulfate; GLS, glucosinolate

\* Corresponding author. University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark.

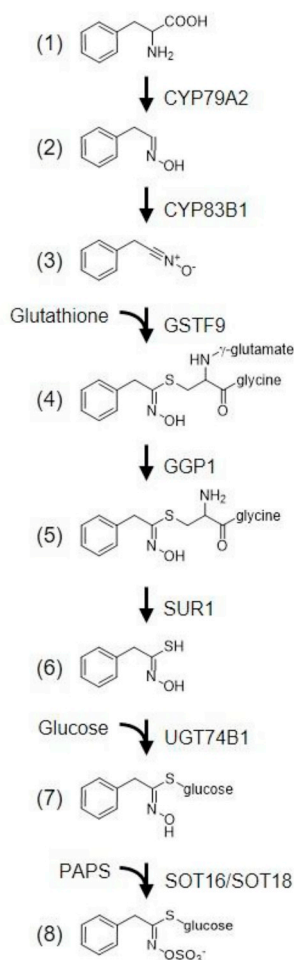
E-mail address: [bah@plen.ku.dk](mailto:bah@plen.ku.dk) (B.A. Halkier).

<https://doi.org/10.1016/j.ymben.2019.02.004>

Received 4 December 2018; Received in revised form 11 February 2019; Accepted 24 February 2019

Available online 01 March 2019

1096-7176/© 2019 The Authors. Published by Elsevier Inc. on behalf of International Metabolic Engineering Society. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



**Fig. 1.** Biosynthetic pathway of benzyl glucosinolate in *Arabidopsis thaliana*. CYP79A2 and CYP83B1 are cytochrome P450s. GSTF9, GGP1, SUR1 and UGT74B1 are a glutathione-S-transferase,  $\gamma$ -glutamyl peptidase, C-S lyase and UDP-glucosyltransferase. The SOT16 and SOT18 are sulfotransferases. The pathway compounds are 1) phenylalanine, 2) phenylacetaldoxime, 3) phenylacetone nitrile oxide, 4) S-[(Z)-phenylacetohydroximoyl]-L-glutathione, 5) Cys-Gly conjugate, 6) phenylacetothiohydroxamic acid, 7) desulfo-benzyl glucosinolate (dsBGLS) and 8) benzyl glucosinolate (BGLS).

elongation of the amino acid side chain before entering the core structure pathway, and the side chains can be further modified as a final step (Sønderby et al., 2010). Expression of multi-gene pathways in microbial hosts can be difficult. The core GLS pathway is further complicated by having two cytochrome P450 enzymes and two sulfur-

incorporating steps. The feasibility of engineering the GLS pathway in a heterologous host was first shown by transient expression of the benzyl GLS (BGLS) genes in *Nicotiana benthamiana* (Geu-Flores et al., 2009b). This work identified an additional enzyme - a  $\gamma$ -glutamyl peptidase, GGP1 - in the pathway (Geu-Flores et al., 2009a), and showed that the reduced sulfur atom was derived from glutathione (Geu-Flores et al., 2011). The final step of the pathway - the conversion of desulfo-BGLS (dsBGLS) to BGLS - was identified as a metabolic bottleneck. This conversion is catalysed by a sulfotransferase, which requires 3'-phosphoadenosine-5'-phosphosulfate (PAPS) as co-factor. The metabolic bottleneck was alleviated by co-expressing the *Arabidopsis* kinase APK2 responsible for PAPS production (Møldrup et al., 2011). Having identified and overcome these basic metabolic bottlenecks in *N. benthamiana* provided a basis for engineering of GLSs in microbial hosts. The tryptophan-derived GLS indol-3-ylmethyl GLS (I3M) was produced at 1.07 mg/L in *S. cerevisiae* proving for the first time that GLSs can be produced in a microbial organism (Mikkelsen et al., 2012). Recently, two reports of engineering of GLS biosynthetic genes in *E. coli* have been published. In 2016, the chain elongation pathway for methionine was successfully expressed in *E. coli* resulting in the production of di-homo-methionine, which is precursor to 4-methylsulfinylbutyl GLS (4MSB) (Mirza et al., 2016). In 2018, 4MSB was produced in *E. coli* although production levels were not quantified (Yang et al., 2018).

In this work, we set out to establish microbial production of the phenylalanine-derived BGLS and systematically optimise the yields. *E. coli* was chosen as production host, as previous reports showed it was capable of expressing a GLS pathway despite its complexity (Yang et al., 2018). The choice of expression strain, medium, cultivation conditions and construct designs were all factors shown to have major effects on BGLS yield in our production system. Targeted proteomics was used to monitor the relative protein expression levels. We optimised production levels to 20.3  $\mu$ M (or 8.3 mg/L) BGLS.

## 2. Materials and methods

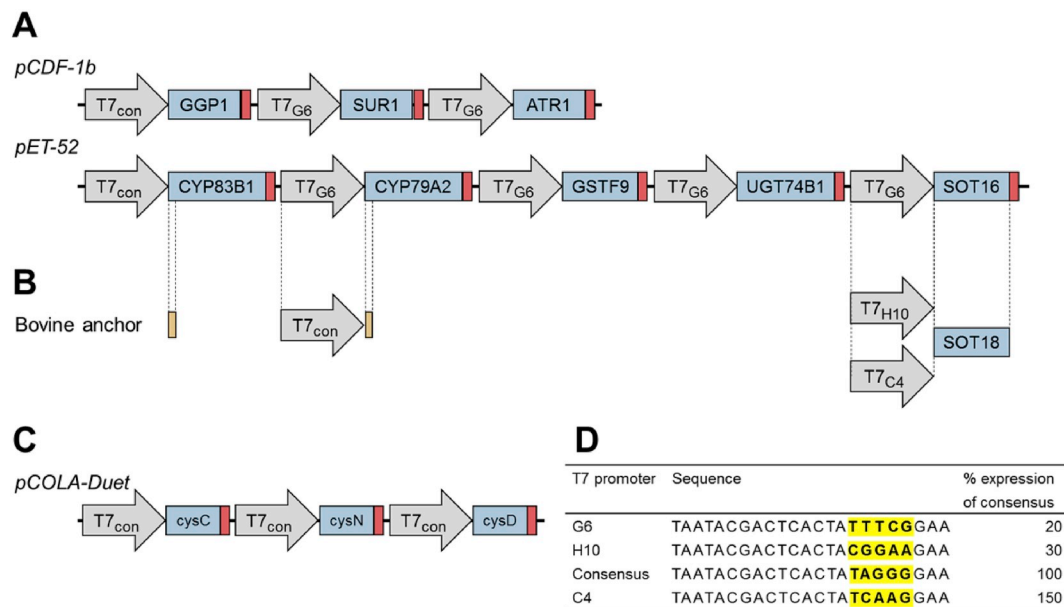
### 2.1. Construct design and generation

The GLS biosynthetic gene sequences were found at [www.arabidopsis.org](http://www.arabidopsis.org) and codon optimised for *E. coli* expression using the Codon Optimisation Tool from IDT (Integrated DNA Technologies Inc, USA <http://eu.idtdna.com/CodonOpt>). The optimised gene sequences were ordered as gBlocks from IDT (Integrated DNA Technologies Inc, USA) and cloned into expression vectors by Genewiz (New Jersey, USA). The *E. coli* sulfate assimilation gene sequences *cysC* (adenylyl-sulfate kinase), *cysN* (subunit of sulfate adenylyltransferase) and *cysD* (subunit of sulfate adenylyltransferase) were found in UniprotKB database and PCR amplified with Phusion® High Fidelity Polymerase (New England Biolabs, USA) from genomic DNA of the BL21 (DE3) *E. coli* strain (Table 1).

**Table 1**

Engineered biosynthetic and support genes for BGLS production in *E. coli*. Gene names are listed with the type of enzyme, genomic locus in native host, length of gene and percentage of nucleotide identity between wildtype and codon-optimised sequence.

Gene	Type	Locus	Length	Sequence identity	Reference
CYP79A2	Cytochrome P450	At5g05260	1572 bp	74.9%	Wittstock and Halkier (2000)
CYP83B1	Cytochrome P450	At4g31500	1500 bp	75.7%	Naur et al. (2003)
GSTF9	Glutathione S-transferase	At2g30860	648 bp	76.2%	Mikkelsen et al. (2012)
GGP1	$\gamma$ -glutamyl peptidase	At4g30530	753 bp	74.8%	Geu-Flores et al. (2009a)
SUR1	C-S lyase	At2g20610	1389 bp	74.7%	Mikkelsen et al. (2004)
UGT74B1	Glycosyltransferase	At1g24100	1383 bp	74.8%	Grubb et al. (2004)
SOT16	Sulfotransferase	At1g74100	1017 bp	76.4%	Piotrowski et al. (2004)
SOT18	Sulfotransferase	At1g74090	1053 bp	76.5%	Piotrowski et al. (2004)
ATR1	P450 reductase	At4g24520	2079 bp	73.8%	Mizutani and Ohta (1998)
<i>cysC</i>	Adenylyl-sulfate kinase	JW2720	606 bp	100%	Satishchandran et al. (1992)
<i>cysN</i>	Sulfate adenylyltransferase, subunit 1	JW2721	1428 bp	100%	Leyh et al. (1988)
<i>cysD</i>	Sulfate adenylyltransferase, subunit 2	JW2722	909 bp	100%	Leyh et al. (1988)



**Fig. 2.** Design of multigene constructs for expression of BGLS pathway and support genes in *E. coli*. Promoters and ribosomal binding site sequences are represented as grey arrows. Biosynthetic genes, terminator sequences and P450 bovine anchor are boxes coloured blue, red and yellow, respectively. A) The two constructs containing the BGLS biosynthetic genes with choice of promoter indicated. B) For optimisation changes to the two constructs included bovine modification of N-terminal anchor of P450 enzymes, change of promoters as well as use of different sulfotransferase genes. C) Construct used to overexpress the *E. coli* PAPS synthesis pathway (adenylyl-sulfate kinase (*cysC*) and the two subunits of sulfate adenylyltransferase (*cysD* and *cysN*)), referred to as *cysCND*. D) T7 promoters used in this study (Table modified from Jones et al. (2015)). Sequence differences are highlighted and the expression levels relative to the consensus T7 promoter are listed. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

For this work pET-52b (Novagen<sup>®</sup>, Merck, #71554), pCDF-1b (Novagen<sup>®</sup>, Merck, #71330) and pCOLADuet<sup>™</sup>-1 (Novagen<sup>®</sup>, Merck, #71406) were used as expression vectors. *GGP1*, *SUR1* and *ATR1* were expressed from pCDF-1b (~20–40 copies/cell). The remaining core structure genes were expressed from pET-52b (~40 copies/cell). The *E. coli* sulfate assimilation pathway genes leading to PAPS, i.e. the two subunits of sulfate adenylyltransferase (*cysD* and *cysN*) and the adenylyl-sulfate kinase (*cysC*) were expressed from pCOLADuet (~20–40 copies/cell). The construct is referred to as *cysCND* (Fig. 2C). All genes were organised in separate operons using four different T7 promoters with varying strengths: consensus, G6, H10 and C4 (naming as in (Jones et al., 2015)). In the constructs indicated, the first eight residues of the N-terminus anchors on the two P450 enzymes were exchanged with a modified anchor of bovine CYP17A1 (hereafter bovine-modified) optimised for *E. coli* expression (Barnes et al., 1991). DNA fragments encoding for the modified P450 anchor, the *SOT18* enzyme as well as the promoters for *CYP79A2* and *SOT* enzymes were exchanged by ligation cloning with T4 DNA ligase (Fig. 2 and Supplemental Fig. S1). The sequences were ordered as gBlocks with appropriate restriction sites (Integrated DNA Technologies Inc, USA) and ligated into the expression vectors.

**Table 2**

Expression strains used for BGLS production. The strains are listed with the name used throughout this work and with strain specific characteristics.

Strain	Name used	Characteristics	Reference
BL21 (DE3)	BL21 (DE3)	Genome integrated <i>placUV5::T7</i> RNA polymerase	NEB <sup>®</sup> C25271
BL21 (DE3) pLysS	pLysS	BL21 (DE3) containing plasmid-expressed T7 lysozyme	Promega L1191
Rosetta <sup>™</sup> (DE3)pLysS	Rosetta	pLysS including rare tRNAs on plasmid	Novagen <sup>®</sup> , Merck 70956
BL21 (DE3) with pRI952	pRI952	BL21 (DE3) with rare tRNAs on plasmid	Del Tito et al. (1995)
One Shot <sup>®</sup> BL21-AI <sup>™</sup>	BL21-AI	Genome integrated <i>paraBAD::T7</i> RNA polymerase	ThermoFisher C6070-03
NEB <sup>®</sup> 5-alpha with pTARA <sup>a</sup>	DH5a	Cloning strain with plasmid-expressed <i>paraBAD::T7</i> RNA polymerase	NEB <sup>®</sup> C2987H
NEB <sup>®</sup> 10-beta with pTARA <sup>a</sup>	DH10B	Cloning strain with plasmid-expressed <i>paraBAD::T7</i> RNA polymerase	NEB <sup>®</sup> C3019H

<sup>a</sup> pTARA plasmid constructed in Wycuff and Matthews (2000).

## 2.2. *E. coli*: strains, growth conditions and transformation

All cloning and plasmid amplifications were done in NEB<sup>®</sup> DH10B strain (New England Biolabs, #C3019H). Cells were cultivated in Luria-Bertani broth (LB Lennox, Duchefa, #L1703) containing 10 g tryptone, 5 g yeast extract and 5 g sodium chloride per litre, or Terrific Broth (TB) containing 12 g tryptone, 24 g yeast extract, 5% glycerol (w/v), 0.17 M KH<sub>2</sub>PO<sub>4</sub> and 0.72 M K<sub>2</sub>HPO<sub>4</sub> per litre. Transformations were performed by heat-shock and the cells recovered in LB at 37 °C for 60 min before plating on LB agar plates with appropriate selection (50 µg/ml carbenicillin, spectinomycin or kanamycin).

The constructs containing the core structure genes were transformed into seven different expression strains (Table 2): four different BL21 (DE3) strains inducible by isopropyl β-D-thiogalactopyranoside (IPTG), as well as one BL21-AI strain and two cloning strains with the T7 RNA polymerase expressed from a plasmid (pTARA, (Wycuff and Matthews, 2000)) and inducible by arabinose. Five of the strains contain an extra plasmid (pLysS, Rosetta, pRI952 or pTARA), which is compatible with co-expression of the three expression plasmids (pET-52b, pCDF-1b and pCOLA-duet).

For expression cultures, LB media was inoculated with single

colonies and incubated over night at 37 °C, 200 rpm. The expression cultures were cultivated in LB or TB as indicated and with the appropriate antibiotics (50 µg/mL carbenicillin, spectinomycin and/or kanamycin, 15 µg/mL chloramphenicol). Experiments were performed in 24-well plates with 4 mL culture. Fresh media with appropriate selection was inoculated to an OD<sub>600</sub> of approximately 0.1 with the overnight culture and allowed to grow at 37 °C and 200 rpm to an OD<sub>600</sub> of approximately 0.6. At this point the cultures were induced with 0.5 mM IPTG or 0.1% (w/v) arabinose and grown at 18 °C and 120 rpm for 24 or 48 h before harvesting. Three individual colonies from each strain were tested in triplicates for the metabolite analysis. When the cultures were supplemented with various intermediates, the following concentrations were added: 25 µM phenylacetaldoxime, 75 mg/L 5-aminolevulinic acid, 20 µM glutathione, 20 µM cysteine, 1 mM thiamine or 20 µM magnesium sulfate. When the strain overexpressing the *cysCND* construct was supplemented with sulfur either 1 mM cysteine and/or 50 µM magnesium sulfate were used.

### 2.3. Metabolite extraction and LC-MS analysis

The supernatant was collected after centrifugation at 13,000 × g for 5 min. An aliquot of the supernatant was diluted 10-fold with milliQ grade water and finally 1:10 (v/v) mixed with a stock solution containing 10 µg/mL 13C-, 15N-labelled amino acids (Algal amino acids 13C, 15N, Isotec, Miamisburg, US) and 25 µM sinigrin (PhytoLab, Vestenbergsgreuth, Germany). The resulting 100-fold diluted samples were filtered (Durapore® 0.22 µm PVDF filters, Merck Millipore, Tullagreen, Ireland) and used directly for LC-MS analysis.

The LC-MS analysis was performed as previously described (Mirza et al., 2016) with changes as detailed below to monitor amino acid levels simultaneously with products and intermediates of the BGLS biosynthesis pathway. Briefly, chromatography was performed on an Advance UHPLC system (Bruker, Bremen, Germany). Separation was achieved on a Zorbax Eclipse XDB-C18 column (100 × 3.0 mm, 1.8 µm, Agilent Technologies, Germany). Formic acid (0.05% (v/v)) in water and acetonitrile (supplied with 0.05% (v/v) formic acid) were employed as mobile phases A and B, respectively. The elution profile was: 0–1.2 min 3% B; 1.2–4.3 min 3–65% B; 4.3–4.4 min 65–100% B; 4.4–4.9 min 100% B, 4.9–5.0 min 100–3% B and 5.0–6.0 min 3% B. Mobile phase flow rate was 500 µL/min and column temperature was maintained at 40 °C. The liquid chromatography was coupled to an EVOQ Elite TripleQuad mass spectrometer (Bruker, Bremen, Germany) equipped with an electrospray ionisation source (ESI). Instrument parameters were optimised by infusion experiments with pure standards. The ionspray voltage was maintained at 3000 V or –4000 V in positive or negative ionisation mode, respectively. Cone temperature was set to 300 °C and cone gas flow to 20 psi. Heated probe temperature was set to 400 °C and probe gas flow set to 50 psi. Nebulising gas was set to 60 psi and collision gas to 1.6 mTorr. Nitrogen was used as both cone gas and nebulising gas and argon as collision gas.

Multiple reaction monitoring (MRM) was used to monitor analyte parent ion → product ion transitions: MRM for the 13C-, 15N-labelled phenylalanine was chosen as previously described (Docimo et al., 2012). MRMs for dsBGLS, BGLS and sinigrin were chosen as previously described (Crocchi et al., 2016). Both Q1 and Q3 quadrupoles were maintained at unit resolution. Bruker MS Workstation software (Version 8.2.1, Bruker, Bremen, Germany) was used for data acquisition and processing. BGLS and dsBGLS were quantified using sinigrin and 13C, 15N-Phe, respectively, as internal standards, except for Fig. 3A, where quantification was performed based on a dilution series (linear regression). Response factors were calculated based on dilution series of the respective analytes. Internal standards were chosen based on matching ionisation mode with the analyte of interest (i.e. negative ionisation mode for glucosinolates and positive ionisation mode for desulfo-glucosinolates and amino acids). Matrix effects from media and cultivating *E. coli* were found to impair correct quantification of the

analytes. The correction factors for matrix effects have been calculated into the response factors given in Supplemental Table S1. The presence of the phenylacetaldoxime and the GSH-conjugated intermediate were monitored without quantification. For analytes where multiple transitions were monitored, the transition used for quantification is marked as quantifier (Qt). Further details for transitions and collision energies can be found in Supplemental Table S1. Visualisation and statistical analysis were performed in RStudio v1.0.153 (R version 3.4.1) (RStudio Team, 2015) using (Garnier, 2018; Wickham, 2016, 2011, 2007).

### 2.4. Targeted proteomics

#### 2.4.1. Design of reference peptides

For each protein of interest, a set of proteotypic peptides was designed and ordered as stable isotopically labelled synthetic peptides (JPT, SpikeTides™). These were used as internal references in the samples. Recently, a set of peptides covering most of the BGLS pathway was designed and generously provided to us by Meike Burow and Daniel Vik. We designed peptides for the remaining enzymes of the pathway (*ATR1* and *GSTF9*) and housekeeping enzymes from *E. coli*. A list of peptides was obtained from each protein through an *in silico* tryptic digest in Skyline 4.2 (MacLean et al., 2010). Suitable peptides were selected based on the following criteria: 1) less than 1250 *m/z*, 2) no neighbouring arginine or lysine residue to the cleavage site, and 3) avoidance of methionine and cysteine residues if possible. Each enzyme was covered by a minimum of two non-neighbouring peptides (Table 3).

#### 2.4.2. Protein extraction and tryptic digest

Pellets from 1.5 mL culture were harvested by centrifugation and kept at –20 °C until extraction. Total protein extraction was performed as previously described (Wessel and Flügge, 1984) with one exception: centrifugation steps were done at 21,000 × g for 1 min at 4 °C. The protein pellet was dried in Speed-Vac (30–60 min, 1000 rpm, max 35 °C). The downstream protocol was modified from previously described methods (Batth et al., 2012). The protein pellet was extracted twice in 100 µL 100 mM ammonium bicarbonate buffer with 10% (v/v) methanol by sonicating for 2–5 min in ultrasonic bath followed by a

**Table 3**

Proteotypic peptides of BGLS biosynthetic enzymes and an *E. coli* housekeeping enzyme used for targeted proteomics analysis. Isotopically labelled synthetic peptides were used as internal reference. Peptides marked with asterisk were used for relative quantification.

Enzyme	AGI code	Peptide sequence	Reference
CYP79A2	At5g05260	SWPLIGNLPEILGR	This study
		GNAFVVIDLR	This study
		LVIESDLPNLNYVK	This study
		LIQGFVWLPVPGK*	This study
CYP83B1	At4g31500	GYVSEEDIPNLPLYK*	This study
		GQDFELLFPFGSGR	This study
		LAVISSAELAK	This study
		LAGVLDVVEAHLK*	This study
GSTF9	At2g30860	GVAFETIPVDLMK	This study
		DAITPGSYFGNEIPDSIAIK*	This study
GGP1	At4g30530	VVSGEFPDEK	This study
		FASIVPVLTLAGISK	This study
SUR1	At2g20610	IGWIALNDPEGVFETTK*	This study
		EENLVFLPGDALGLK	This study
UGT74B1	At1g24100	GLPSLSYDELPSFVGR	This study
		SINEFIESLKG*	This study
SOT16	At1g74100	VGDWANYLTPEMAAR	This study
SOT18	At1g74090	YDDAANPLLK*	This study
		FDDSSNPLLK	This study
ATR1	At4g24520	VVNLCSEFTLK*	This study
		VIDLDDYAADDQYEEK*	This study
ICD	JW1122	DEDDDLDLGSGK	This study
		GPLTTPVGGGIR	Batth et al. (2014)



brief centrifugation (1.5 min at  $1,000\times g$ ). Supernatant of both extractions was combined and protein concentration determined by Pierce™ BCA Protein Assay Kit (ThermoFisher, #23225). For tryptic digest, 100  $\mu$ l of 0.5 mg/ml protein extraction was used (50  $\mu$ g total protein). The extraction was incubated with 1  $\mu$ l 10 mM DTT for 30 min at room temperature. Subsequently, 25  $\mu$ l 50 mM iodoacetamide was added and the samples incubated in the dark for 20 min at room temperature. An additional 100  $\mu$ l 100 mM ammonium bicarbonate buffer with 10% (v/v) methanol was added with 1  $\mu$ g trypsin/Lys-C mix (Promega, #V5073) and the samples incubated over night at 37 °C. The digest was stopped by adding 20  $\mu$ l 10% (v/v) trifluoroacetic acid (TFA) and samples were diluted up to 1.5 ml in buffer A (2% (v/v) acetonitrile and 0.1% (v/v) formic acid) and centrifuged (20,000  $\times g$  for 15 min). Peptides were purified over Sep-Pak C-18 columns (Waters, Sep-Pak® Vac 1 cc 100 mg, #WAT023590). Columns were attached to vacuum manifold and equilibrated with 1 ml buffer B (65% (v/v) acetonitrile and 0.1% (v/v) formic acid) followed by 1 ml buffer A. Next, the peptide digest was loaded, followed by three washes with 1 ml buffer A and finally eluted with two times 0.5 ml buffer B. Purified peptides were dried by Speed-Vac (2–4 h, 1000 rpm, max 35 °C). Dried peptides were kept at –20 °C until analysis. Just prior to LC-MS analysis, the dried peptides were resuspended in 25  $\mu$ l buffer C (2% (v/v) acetonitrile, 0.5% (v/v) formic acid and 0.1% (v/v) TFA), spiked with 20 nM isotopically labelled peptide standards (JPT, SpikeTides™), and filtered through 0.22  $\mu$ m centrifugal filters (#UFC30GV00, Merck, Darmstadt, Germany).

#### 2.4.3. LC-MS method

The gradient was adopted from Batth et al. (2014) and Percy et al. (2012) with modifications. Briefly, formic acid (0.1% (v/v)) in water and acetonitrile (supplied with 0.1% (v/v) formic acid) were employed as mobile phases A and B, respectively. The elution profile was: 0–1.0 min 5–10% B; 1.0–3.0 min 10–11% B; 3.0–13.0 min 11–19% B; 13.0–21.0 min 19–27.5% B; 21.0–21.7 min 27.5–24% B; 21.7–22.5 min 34–42% B; 22.5–23.5 min 42–90% B; 23.5–26.9 min 90% B; 26.9–30.0 min 90–5% B and 30.0–34.0 min 5% B. Mobile phase flow rate was 500  $\mu$ L/min and column temperature was maintained at 55 °C. Peptide separation was achieved on an Aeris PEPTIDE, XB-C18 column (1.7  $\mu$ m, 2.1  $\times$  150 mm, Phenomenex, Palo Alto, USA) on an Advance UHPLC-OLE (Bruker Daltonics, Bremen, Germany). The injection volume was 10  $\mu$ L.

Source settings for heated electrospray ionisation were as follows: spray voltage 3200 V, in positive ionisation mode; cone temperature 300 °C; cone gas flow 20 psi; heated probe temperature 300 °C; probe gas flow 40 and nebuliser gas flow 50. Nitrogen was used as cone and probe gas and argon as collision gas. The triple quadrupole mass spectrometer (EVOQ Elite, Bruker Daltonics, Bremen, Germany) was set to scan for parent ion  $\rightarrow$  product ion transitions for individual peptides within scheduled 3 min windows. Resolution of the first and third quadrupole was set to  $\pm 1$  Da. Detailed information on peptides including retention times, transitions selected for detection and quantification and collision energies are summarised in Supplemental Table S2. The acquired chromatograms were manually inspected and ratios between endogenous light and synthetic heavy peptides were obtained through Skyline 4.2 for each peptide (MacLean et al., 2010). Relative quantification was performed by normalisation to heavy peptides, followed by normalisation to the housekeeping protein ICD and plotted relative to the CSaro control. Undetected peptides (N/A) were treated as zero values for analysis. Additional statistical analyses and visualisation were performed in RStudio v1.0.153 (R version 3.4.1) (RStudio Team, 2015) using (de Mendiburu, 2017; Wickham, 2016, 2011, 2007).

### 3. Results

#### 3.1. Establishment of dsBGLS production in *E. coli*

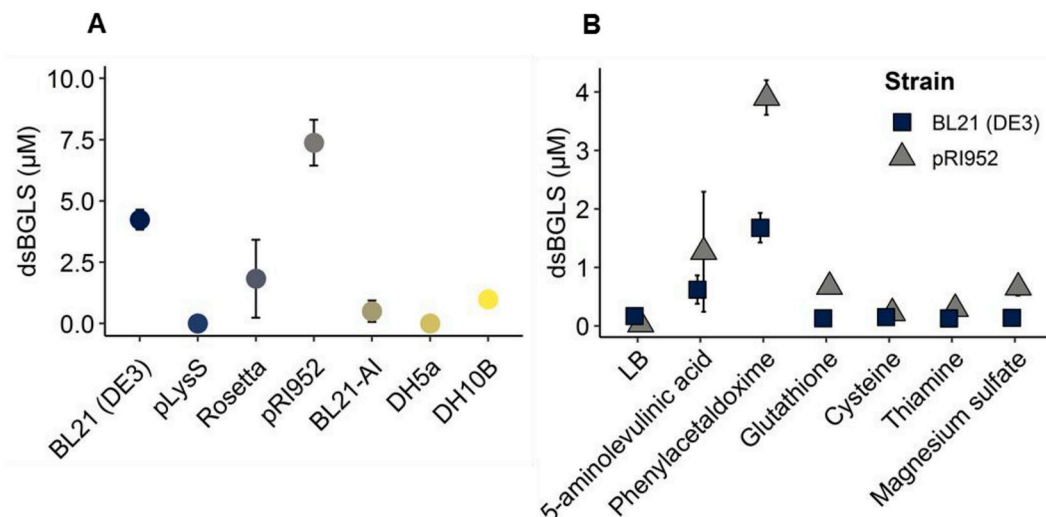
The biosynthetic pathway of phenylalanine-derived BGLS consisting of seven genes - *CYP79A2*, *CYP83B1*, *GSTF9*, *GGP1*, *SUR1*, *UGT74B1* and *SOT16* - was chosen for engineering and optimising GLS production in *E. coli*. The NADPH cytochrome P450 reductase, *ATR1*, from *A. thaliana* was included, as *E. coli* does not efficiently support the function of cytochrome P450 enzymes (Park et al., 2013). The pathway was expressed from plasmids to enable efficient and rapid modifications of the construct design. The genes were divided between two medium to high copy-number plasmids and all genes were expressed from IPTG- or arabinose-inducible T7 promoters (Fig. 2A). Different T7 promoters were used as a previous study showed that choice of T7 promoters can affect production levels substantially (Jones et al., 2015). Based on previous findings on the evolutionarily related cyanogenic glucoside biosynthetic pathway (Laursen et al., 2016) and on the low frequency by which *CYP79s* compared to *CYP83s* are found in classical shotgun proteomics approaches in *Arabidopsis*, we expected *CYP83s* to be present in higher levels than *CYP79s* in planta (Baerenfaller et al., 2011; Laursen et al., 2016). We therefore chose to express *CYP83B1* from a consensus T7 promoter and the other genes from the weaker G6 promoter.

First, the widely used BL21 (DE3) expression strain was engineered with the BGLS pathway. Besides the desired product BGLS, three intermediates of the pathway were monitored: phenylacetaldoxime, the GSH conjugate and dsBGLS (Fig. 1). No product was detected when expression was performed at 28 °C. Upon lowering the temperature to 18 °C, dsBGLS was produced, but still none of the other monitored products were detected (data not shown). As dsBGLS is the final intermediate of the pathway (Fig. 1), this result indicated functionality of all but the ultimate sulfotransferase step. The OD<sub>600</sub> at the time of induction proved important to titres with optimal production after an induction at OD<sub>600</sub>  $\approx$  0.6 (data not shown). Harvesting 48 h after induction resulted in 4.74  $\mu$ M dsBGLS, but still no BGLS was detected. Based on these results, subsequent experiments were performed by inducing expression at OD<sub>600</sub>  $\approx$  0.6 and allowing production cultures to grow at 18 °C for 48 h after induction.

#### 3.2. Screening of different *E. coli* strains for production

The plasmids containing the BGLS genes were transformed into six additional *E. coli* strains to investigate whether strain specificity would result in BGLS production. The BL21 (DE3) strain was included as control. Strains pRI952 and Rosetta are engineered with rare tRNAs to help expression. BL21-AI has the T7 RNA polymerase expressed from the *araBAD* promoter enabling repression of potential toxic, leaky expression by glucose (Miyada et al., 1984). The pLysS strain represses expression until IPTG induction through the T7 lysozyme (Studier, 1991). These five strains are all of the classical B-lineage of *E. coli*. Two K-12-derived strains, DH5a and DH10B, were included as an alternative. These were transformed with the pTARA plasmid carrying the T7 RNA polymerase expressed from the *araBAD* promoter like the BL21-AI strain (Wycuff and Matthews, 2000).

When seven different *E. coli* strains harbouring the BGLS biosynthetic genes were grown for 48 h at 18 °C, differences in dsBGLS production were observed (Fig. 3A). The highest-producing strains were the standard strain BL21 (DE3) and the pRI952. The latter produced 1.6-fold more dsBGLS than BL21 (DE3). These two strains were selected for further optimisations.



**Fig. 3.** Production level of dsBGLS in LB medium of *E. coli* expressing the BGLS biosynthetic genes. A) Seven different *E. coli* strains were grown for 48 h at 18 °C and the media were analysed for GLS production. B) The BL21 (DE3) and pRI952 strains were grown in media supplemented with either 25 μM phenylacetaldoxime, 75 mg/L 5-aminolevulinic acid or different sulfur sources (20 μM glutathione, 20 μM cysteine, 1 mM thiamine or 20 μM magnesium sulfate). Data represent the average and standard deviation of three biological replicates each grown in three technical replicates.

### 3.3. Optimisation of medium composition

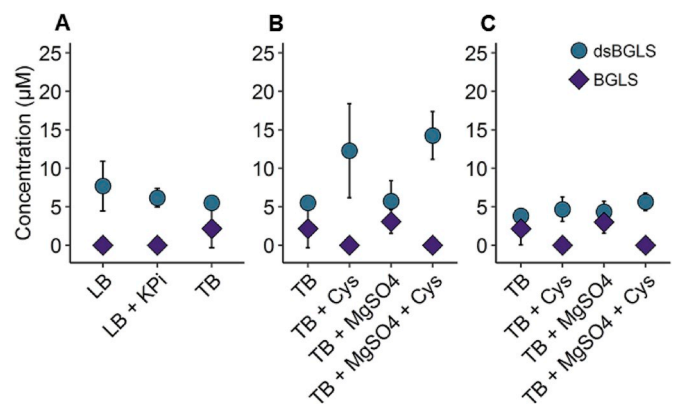
#### 3.3.1. Supplying additives or intermediates of the pathway

The first two enzymes of the BGLS pathway are the heme-containing P450 proteins. Supplementation of medium with heme precursor 5-aminolevulinic acid was previously shown to improve activity of P450 enzymes in *E. coli* (Hansen et al., 2001; Pritchard et al., 1997; Wittstock and Halkier, 2000). When the expression cultures were fed with 5-aminolevulinic acid to investigate whether co-factor supplement would result in BGLS production only little increase in dsBGLS levels and still no BGLS was observed (Fig. 3B). When feeding the first intermediate of the BGLS pathway, phenylacetaldoxime, production of dsBGLS was increased 10-fold and 7-fold in strains BL21 (DE3) and pRI952, respectively. This indicated that the entry point to the pathway was a metabolic bottleneck.

The BGLS pathway contains two sulfur-incorporating enzymatic steps; the glutathione-conjugation reaction by GSTF9 and the transfer of sulfate by SOT16 (Fig. 1). When the cultures were supplemented with glutathione, no effect on production levels was observed suggesting that glutathione is not limiting (Fig. 3B). The second sulfur-incorporation step is responsible for the conversion of dsBGLS to BGLS, and hence the step of the pathway, which appeared non-functional. Sulfotransferase (SOT) enzymes transfer a sulfate moiety from PAPS onto dsGLSs (Klein and Papenbrock, 2009). PAPS is a product of the sulfate assimilation pathway (Kopriva and Koprivova, 2004). To increase the SOT activity and obtain BGLS, different sulfur sources were fed to the expression cultures in the form of cysteine, thiamine and magnesium sulfate. This did not result in any BGLS production, and both cysteine and thiamine appeared to hamper production of dsBGLS (Fig. 3B). The pRI952 strain proved consistently better than the BL21 (DE3) and subsequent experiments were carried out using only this strain.

#### 3.3.2. Production of BGLS using TB medium

Instead of supplementing individual compounds to the medium, the LB medium was exchanged with the more nutritious TB medium that predominantly differs by containing approximately 5-fold more yeast extract and by being buffered by potassium phosphate (KPi). BGLS was produced in TB to levels of approximately 4.1 μM (Fig. 4A). This is comparable to the production of I3M in *S. cerevisiae* (2.38 μM, Mikkelsen et al., 2012). Only dsBGLS was detected in LB medium. When KPi buffer was added to LB medium to investigate if the



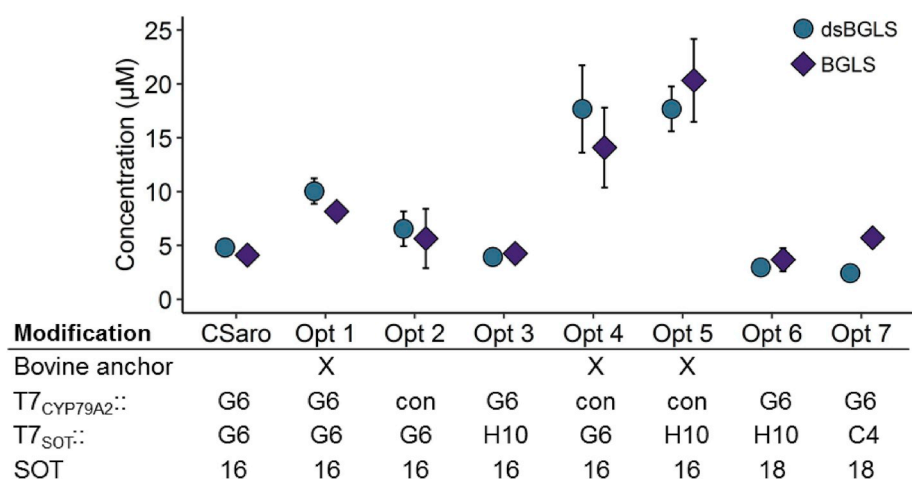
**Fig. 4.** Medium levels of dsBGLS and BGLS produced by pRI952 *E. coli* strain grown in different media and with sulfur supplements. A) Production of dsBGLS and BGLS in medium by the pRI952 strain grown in LB, LB with KPi buffer or TB medium. B) as A) except the TB medium is supplemented with different sulfur sources. C) as B) except the strain harbours the cysCND construct in addition to the BGLS biosynthesis genes. Data represent the average and standard deviation of three biological replicates each grown in three technical replicates.

phosphate buffer and thus the control of pH was the determining factor for dsBGLS to BGLS conversion in TB medium, still only dsBGLS was produced. This suggests that component(s) in the yeast extract are critical for SOT activity (Fig. 4A).

### 3.4. Towards elimination of metabolic bottlenecks

#### 3.4.1. Overexpression of *E. coli* PAPS-producing genes

Approximately one third of the produced dsBGLS is converted to the desired product, BGLS, when cultivating the pRI952 strain in TB medium. As mentioned above, the sulfotransferase responsible for this conversion needs PAPS as a co-factor. In *E. coli*, the genes responsible for producing PAPS in the sulfate assimilatory pathway are *cysC* encoding for an adenylyl-sulfate kinase and *cysN* and *cysD* encoding for two subunits of a sulfate adenylyltransferase (Kopriva and Koprivova, 2004). To increase the levels of available PAPS, a construct harbouring these genes driven by different T7 promoters was generated (Fig. 2C, construct cysCND). When cultures of pRI952 strain expressing the BGLS pathway genes and overexpressing cysCND were supplemented with



**Fig. 5.** Production of dsBGLS and BGLS in medium from pRI952 *E. coli* strain upon optimisation of CYP and SOT activity. The two P450 enzymes were N-terminally bovine-modified and CYP79A2 expressed from a stronger T7 promoter. Both SOT16 and SOT18 were tested driven from different T7 promoters. Table indicates the changes made to the original constructs containing the BGLS biosynthetic genes (CSaro). Data represent the average and standard deviation of three biological replicates each grown in three technical replicates. Pairwise *t*-tests were performed on the dataset. Results can be found in [Supplemental Tables 3–5](#).

magnesium sulfate or cysteine or both as sulfur sources, no increase in the BGLS levels was observed, on the contrary BGLS levels were reduced (Fig. 4B and C). Similarly, in the strain expressing only the BGLS pathway genes and fed with added cysteine, BGLS levels were also reduced, showing that BGLS production is reduced when cysteine is supplemented, despite overexpression of the PAPS producing genes. However, supplementing with cysteine more than doubled the production of dsBGLS by the wildtype strain compared to the *cysCND*-overexpressing strain (Fig. 4B and C). This could possibly be explained by higher levels of glutathione as a result of endogenous regulation to establish cellular equilibrium. This regulation may have been disrupted in strains expressing the *cysCND* construct, which could explain the lack of increased dsBGLS production (Fig. 4C).

### 3.4.2. Effect of choice of promoters and P450 N-terminal anchors

The increase in dsBGLS upon feeding with the first intermediate of the pathway, phenylacetaldoxime, shows that the CYP79A2 activity is limiting (Fig. 3B), and the general accumulation of dsBGLS rather than BGLS shows that the SOT activity is limiting (Figs. 4 and 5). A series of new constructs were designed as an alternative approach to increase P450 and SOT activity (Fig. 2B). As a means to increase P450 enzyme activity, the anchors on both P450 enzymes were replaced with an N-terminal sequence of the bovine CYP17A1 (Optimisation strategy (hereafter Opt 1) (Barnes et al., 1991). This approach has previously been used successfully to increase the expression of cytochrome P450 enzymes in *E. coli* (Barnes et al., 1991; Goder and Spiess, 2003; Ichinose and Wariishi, 2013), including CYP79A2 (Wittstock and Halkier, 2000). To increase expression of specifically CYP79A2, this gene was placed in front of a consensus T7 promoter (Opt 2), which is stronger than the replaced G6 (Fig. 2D). A similar approach was employed for the SOT16 gene, which was driven by the H10 promoter (Opt 3). These three optimisation strategies were combined into, respectively, Opt 4 (Opt 1 + 2) and Opt 5 (Opt 1 + 2 + 3). Additionally, two constructs were made to optimise the conversion of dsBGLS to BGLS. *A. thaliana* contains three SOT enzymes (SOT16–18) affiliated with the GLS pathways (Piotrowski et al., 2004). SOT16 and SOT18 have been shown to convert dsBGLS to BGLS equally well (Møldrup et al., 2011). To test if SOT18 will perform better in *E. coli*, we replaced SOT16 with SOT18 expressed from both the H10 promoter (Opt 6) and from the stronger C4 promoter (Opt 7).

The highest production levels of dsBGLS and BGLS were seen in strains harbouring a construct containing the following changes: bovine-modified anchors on both CYPs, stronger promoter (consensus T7) for CYP79A2 and stronger promoter (H10) for SOT16 (Opt 5). This combination increased production of dsBGLS and BGLS by 3.7- and 5-fold, respectively, compared to the unmodified construct. Production in the Opt 4 (bovine-modified P450 anchors and stronger promoter for

CYP79A2) was similar to Opt 5, however the conversion of dsBGLS to BGLS was significantly improved in Opt 5, resulting in higher BGLS production when compared to Opt 4. The increased production in Opt 4 and Opt 5, which both have the bovine-modified P450 anchors and a stronger promoter driving CYP79A2, when compared to the remaining constructs indicates that CYP79A2 was a bottleneck in the pathway. This limitation was improved by protein engineering and increased promoter strength for the gene.

The other bottleneck resulting in dsBGLS accumulation and thus involving the SOT enzyme was addressed with a similar strategy. Increasing the strength of the promoter driving SOT16 expression (Opt 3) did improve the conversion of dsBGLS to BGLS when compared to the unmodified construct. In strains expressing SOT18 from the H10 (Opt 6), the conversion of dsBGLS to BGLS is significantly improved compared to the unmodified construct, but not compared to Opt 3. The construct with highest SOT18 expression (Opt 7) also resulted in the most favourable dsBGLS/BGLS ratio (Fig. 5).

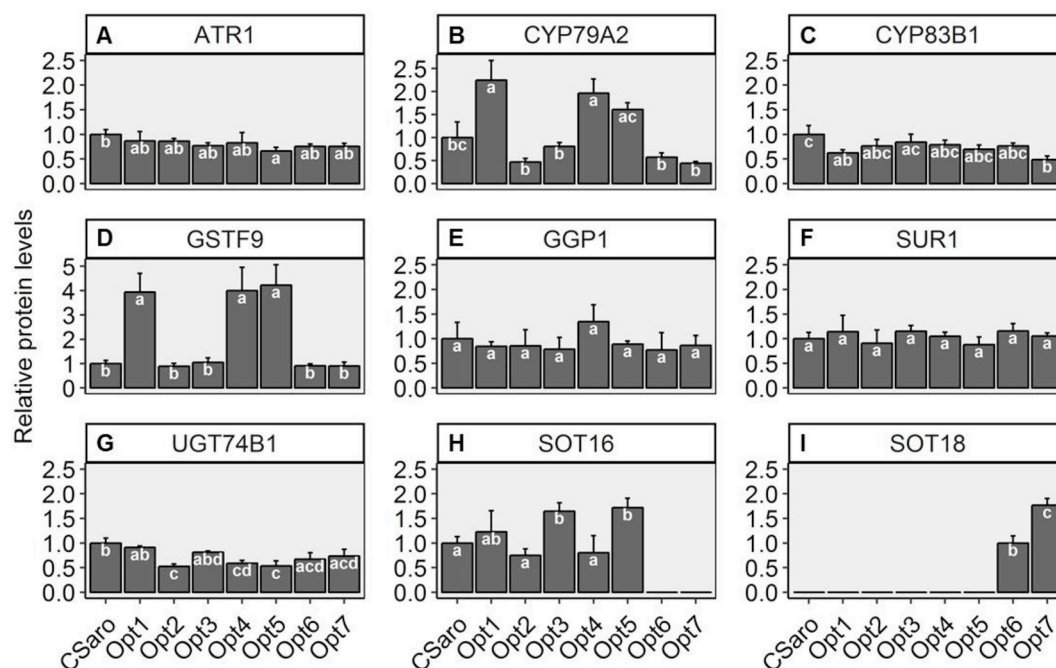
### 3.4.3. Protein expression levels of BGLS biosynthetic enzymes after optimisation

A targeted proteomics approach was applied to monitor effects of the different optimisation strategies on the protein expression levels. For all proteins, the level was determined from a minimum of two peptides, with the exception of GGP1 for which only one peptide was quantifiable (Fig. 6 and Supplemental Fig. S2). Higher levels of CYP79A2 but not CYP83B1 were detected when the N-terminal anchors were bovine-modified (Fig. 6B and C, Opt 1, 4 & 5), but not when the stronger consensus T7 promoter was used for CYP79A2 (Fig. 6B, Opt 2). When the SOT16 and SOT18 were expressed from H10 and C4 promoters, it resulted in increased SOT levels when compared to expression from the G6 promoter (Fig. 6H and I, Opt 3, 6 and 7). GGP1, SUR1 and ATR1 showed similar protein levels in all samples (Fig. 6A, E & F). Unexpectedly, modifying the anchors on the two P450 enzymes resulted in increased GSTF9 levels, which is unchanged at all other conditions (Fig. 6D, Opt 1, 4, & 5).

## 4. Discussion

In this work, we report the first production of BGLS in a microbial host and the first quantified GLS production in *E. coli*. We found that growing the cultures at 18 °C rather than 28 °C and having a cell density of OD<sub>600</sub> at 0.6 at the time of induction was crucial to obtain production. These factors are known to influence production from heterologous pathways and can vary greatly for different compounds (Galloway et al., 2003; Huyen et al., 2015; San-Miguel et al., 2013; Sørensen and Mortensen, 2005). Previous reports have shown that expression temperatures below 20 °C improved production of





**Fig. 6.** Protein levels in pRI952 *E. coli* strain upon optimisation of CYP and SOT activity. Targeted proteomics was used to monitor the protein level of enzymes in BGLS biosynthesis. A representative peptide for each protein is shown. Data represent the average and standard deviation of protein level in three biological replicates normalised to a housekeeping protein and relative to expression seen from the original, unmodified constructs (CSaro). Letters denote significant differences between samples ( $P < 0.05$  Tukey HSD test).

phenylacetaldoxime by CYP79A2 (Miki and Asano, 2014), which corresponds well with our results. Additionally, harvesting after 48 h compared to 24 h as well as choosing the pRI952 strain compared to a standard BL21 (DE3) (Fig. 3A) improved our production 21.5- and 1.6-fold, respectively. The most crucial optimisation for BGLS production was changing to a more nutritious medium composition. In standard LB medium only dsBGLS was produced but upon cultivation of the *E. coli* strains harbouring the BGLS biosynthetic pathway in TB medium, approximately one third of the dsBGLS was converted to BGLS (Fig. 4A). Both, dsBGLS and BGLS were detected in the medium without first disrupting the cells. This corresponds well to previously published microbial GLS engineering studies, which all found the products extracellularly (Mikkelsen et al., 2012; Mirza et al., 2016; Yang et al., 2018).

Metabolic bottlenecks identified in this study are associated with the P450 and SOT enzymes. This is based on an increased production when feeding the CYP79A2 product (phenylacetaldoxime) and optimising the CYP79A2 enzyme expression, as well as improved dsBGLS/BGLS ratio when optimising the SOT enzymes expression. Many studies have focused on optimising activity of eukaryotic P450 enzymes in *E. coli*. Being membrane-associated proteins, overexpression can cause severe growth inhibition and influence the energy metabolism of *E. coli* (Wagner et al., 2007). Particularly, plant P450 enzymes can have problems with proper folding, membrane translocation and toxicity to the host cell (Chemler and Koffas, 2008). In this work, the BGLS-producing *E. coli* strains were not affected in growth compared to strains with empty vector control, and both CYP79A2 and CYP83B1 - representing the entry point to the engineered pathway - were functionally expressed. When the N-terminals on CYP79A2 and CYP83B1 were modified with the bovine anchor known to improve P450 expression in *E. coli* (Barnes et al., 1991; Goder and Spiess, 2003; Ichinose and Wariishi, 2013; Miki and Asano, 2014), protein levels of CYP79A2, but not CYP83B1, increased (Fig. 6B and C) along with the production of both dsBGLS and BGLS (Fig. 5), reflecting increased flux through the pathway. This modification also increased levels of GSTF9 (Fig. 6D) that is expressed just downstream of CYP79A2 in the construct. This apparent linked co-expression between CYP79A2 and GSTF9 could be

explained by an increased transcription of CYP79A2, which mistakenly continues as a read-through to the neighbouring GSTF9. This is, however, unlikely as the nucleotide sequences are identical apart from the modified CYP anchor and the increased protein levels of GSTF9 might be the result of unknown cellular regulatory mechanisms.

For the metabolic bottleneck associated with the final PAPS-dependent sulfotransferase step converting dsBGLS to BGLS, even the best-producing strain converted little more than half of the dsBGLS to BGLS. As the conversion requires incorporation of sulfate, BGLS production likely affects sulfur metabolism of the host. In *N. benthamiana*, the same step was identified as a metabolic bottleneck in BGLS production, but could be alleviated by co-expressing an *Arabidopsis* PAPS kinase, APK2 (Møldrup et al., 2011). Overexpression of *E. coli* adenylyl-sulfate kinase (*cysC*) and sulfate adenylyltransferase (*cysND*) designed to increase the PAPS pool, did not improve conversion of dsBGLS to BGLS (Fig. 4C). In fact, overexpression of these genes appears to be a disadvantage when the cultures are supplemented with additional cysteine. Although BGLS levels were reduced at any level of *cysCND* expression when cysteine is supplemented, dsBGLS was produced in approximately twice as high amounts in wildtype strains compared to *cysCND*-overexpressor strains (Fig. 4B and C). In the former, the additional cysteine may increase the availability of glutathione that is required for the earlier sulfur-incorporating step leading to the thio-glucose moiety in the GLS structure. In the *cysCND*-overexpressor strain regulation of the entire sulfate assimilation pathway may be compromised as most of the pathway is regulated at a transcriptional level (Rossi et al., 2014). This in turn might hamper the cells ability to channel the additional sulfur efficiently towards glutathione and thus dsBGLS production.

Uptake of sulfur from the environment is tightly regulated and dependent on available as well as preferred sources (Guédon and Martin-Verstraete, 2007). For *E. coli*, the main sulfur source from the environment is sulfate, while cysteine if present is the preferred source (Kredich, 2008). Sufficient cysteine in the environment represses expression of the sulfate assimilation pathway. This could well explain the complete loss of BGLS when cysteine was feed to the expression

cultures, as available PAPS would decrease or vanish. Conversely, when cysteine is scarce, expression of the assimilation pathway as well as import of sulfate are increased (Kredich, 2008; Rossi et al., 2014). In the assimilation pathway, both PAPS and its precursor APS (adenosine 5'-phosphosulfate) are signalling molecules. APS signals sufficient inorganic sulfate source and will thus promote the assimilation pathway (Guédon and Martin-Verstraete, 2007). PAPS has broader signalling effects, which are not fully elucidated, but it has been shown to affect global gene transcription through inhibition of nucleotide diphosphate kinase (Ndk) (Rossi et al., 2014). Thus, if overexpression of *cysCND* leads to accumulation of PAPS, this could negatively affect expression of the BGLS pathway genes.

The most promising result with regards to dsBGLS to BGLS conversion was seen in cultures with higher SOT expression. Increasing the SOT enzyme levels by expressing the gene from the strong C4 promoter improved the BGLS/dsBGLS ratio significantly when compared to any other construct. This suggests that the *E. coli* cells contain a PAPS pool, which the SOT enzymes cannot access and/or utilise efficiently.

## 5. Conclusion

Optimisation strategies included test of different expression strains, cultivation conditions and media compositions, as well as modification of flux through the exogenous BGLS pathway and the endogenous sulfur metabolism. A targeted proteomics approach uncovered how changes in genetic regulatory elements affected levels of the BGLS biosynthetic enzymes and showed that increasing promoter strength did not necessarily yield higher protein levels, but could in fact result in decreased levels. Similarly, modifying N-terminal anchors on the P450 enzymes according to previously reported bovine modification had different effects depending on the enzyme in question. Interestingly, increased CYP79A2 levels caused several folds increase in the neighbouring GSTF9. Generally, the modifications to the P450 and SOT genes, which are co-located on the *pET-52b* plasmid, appear to affect the expression levels of other genes expressed from that plasmid. In contrast, genes (*GGP1*, *SUR1*, and *ATR1*) expressed from the *pCDF-1b* plasmid did not alter protein levels under the different optimisation approaches, with the exception of *ATR1* in Opt 5, which is mildly reduced.

The metabolic bottlenecks identified through our work were related to the expression of the two cytochrome P450s and the two sulfur-incorporating steps. The P450 expression was improved by increasing promoter strength for the genes. For the sulfur incorporation, solutions that previously alleviated the same bottleneck in *N. benthamiana*, did not translate into the microbial host system. Feeding sulfur sources increased dsBGLS production but at the cost of BGLS. This increased dsBGLS production indicates that both sulfur-incorporating steps are limiting for the flux through the pathway. Increased expression of the SOT18 protein improved the ratio of dsBGLS/BGLS. However, no condition or strain tested enabled complete conversion of dsBGLS to BGLS. In summary, our work identified and addressed metabolic bottlenecks and optimised the production to 20.3  $\mu$ M BGLS accompanied by 13.5  $\mu$ M dsBGLS.

## Author contributions

All authors took part in designing the study. AP performed all experiments and CC performed metabolite and proteomics LC-MS method development and analysis. All authors wrote the manuscript based on a draft provided by AP.

## Competing financial interests

The authors declare no competing financial interests.

## Acknowledgements

This work was supported by the Danish National Research Foundation [DNRF99]; and the Novo Nordisk Fonden [NNF14OC0011253]. The authors would like to thank Daniel Vik and Meike Burow for providing isotopically labelled peptides for the targeted proteomics analysis. AP thanks the Department of Plant and Environmental Sciences (University of Copenhagen) for a PhD stipend.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ymben.2019.02.004>.

## References

- Alonso-Gutierrez, J., Kim, E.-M., Batth, T.S., Cho, N., Hu, Q., Chan, L.J.G., Petzold, C.J., Hillson, N.J., Adams, P.D., Keasling, J.D., Garcia Martin, H., Lee, T.S., 2015. Principal component analysis of proteomics (PCAP) as a tool to direct metabolic engineering. *Metab. Eng.* 28, 123–133. <https://doi.org/10.1016/j.ymben.2014.11.011>.
- Baerenfaller, K., Hirsch-Hoffmann, M., Svoboda, J., Hull, R., Russenberger, D., Bischof, S., Lu, Q., Gruissem, W., Baginsky, S., 2011. pep2pro: a new tool for comprehensive proteome data analysis to reveal information about organ-specific proteomes in *Arabidopsis thaliana*. *Integr. Biol. (Camb.)* 3, 225–237. <https://doi.org/10.1039/c0ib00078g>.
- Bahadoran, Z., Tohid, M., Nazeri, P., Mehran, M., Azizi, F., Mirmiran, P., 2012. Effect of broccoli sprouts on insulin resistance in type 2 diabetic patients: a randomized double-blind clinical trial. *Int. J. Food Sci. Nutr.* 63, 767–771. <https://doi.org/10.3109/09637486.2012.665043>.
- Barnes, H.J., Arlotto, M.P., Waterman, M.R., 1991. Expression and enzymatic activity of recombinant cytochrome P450 17  $\alpha$ -hydroxylase in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 88, 5597–5601. <https://doi.org/10.1073/pnas.88.13.5597>.
- Batth, T.S., Keasling, J.D., Petzold, C.J., 2012. Targeted proteomics for metabolic pathway optimization. *Methods Mol. Biol.* 944, 237–249. [https://doi.org/10.1007/978-1-62703-122-6\\_17](https://doi.org/10.1007/978-1-62703-122-6_17).
- Batth, T.S., Singh, P., Ramakrishnan, V.R., Sousa, M.M.L., Chan, L.J.G., Tran, H.M., Luning, E.G., Pan, E.H.Y., Vu, K.M., Keasling, J.D., Adams, P.D., Petzold, C.J., 2014. A targeted proteomics toolkit for high-throughput absolute quantification of *Escherichia coli* proteins. *Metab. Eng.* 26, 48–56. <https://doi.org/10.1016/j.ymben.2014.08.004>.
- Brunk, E., George, K.W., Alonso-Gutierrez, J., Thompson, M., Baidoo, E., Wang, G., Petzold, C.J., McCloskey, D., Monk, J., Yang, L., O'Brien, E.J., Batth, T.S., Martin, H.G., Feist, A., Adams, P.D., Keasling, J.D., Palsson, B.O., Lee, T.S., 2016. Characterizing strain variation in engineered *E. coli* using a multi-omics-based workflow. *Cell Syst.* 2, 335–346. <https://doi.org/10.1016/j.cels.2016.04.004>.
- Chae, T.U., Choi, S.Y., Kim, J.W., Ko, Y.-S., Lee, S.Y., 2017. Recent advances in systems metabolic engineering tools and strategies. *Curr. Opin. Biotechnol.* 47, 67–82. <https://doi.org/10.1016/j.copbio.2017.06.007>.
- Chemler, J.A., Koffas, M.A.G., 2008. Metabolic engineering for plant natural product biosynthesis in microbes. *Curr. Opin. Biotechnol.* 19, 597–605. <https://doi.org/10.1016/j.copbio.2008.10.011>.
- Cho, C., Choi, S.Y., Luo, Z.W., Lee, S.Y., 2015. Recent advances in microbial production of fuels and chemicals using tools and strategies of systems metabolic engineering. *Biotechnol. Adv.* 33, 1455–1466. <https://doi.org/10.1016/j.biotechadv.2014.11.006>.
- Chubukov, V., Mukhopadhyay, A., Petzold, C.J., Keasling, J.D., Martín, H.G., 2016. Synthetic and systems biology for microbial production of commodity chemicals. *npj. Syst. Biol. Appl.* 2, 16009. <https://doi.org/10.1038/npjbsa.2016.9>.
- Crocoll, C., Halkier, B.A., Burow, M., 2016. Analysis and quantification of glucosinolates. In: Stacey, G., Birchler, J., Ecker, J., Martin, C.R., Stitt, M., Zhou, J.-M. (Eds.), *Current Protocols in Plant Biology*. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 385–409. <https://doi.org/10.1002/cppb.20027>.
- Dahl, R.H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T.S., Redding-Johanson, A.M., Petzold, C.J., Mukhopadhyay, A., Lee, T.S., Adams, P.D., Keasling, J.D., 2013. Engineering dynamic pathway regulation using stress-response promoters. *Nat. Biotechnol.* 31, 1039–1046. <https://doi.org/10.1038/nbt.2689>.
- de Mendiburu, F., 2017. *Agricolae: Statistical Procedures for Agricultural Research*. Del Tito, B.J., Ward, J.M., Hodgson, J., Gershater, C.J., Edwards, H., Woskoy, L.A., Watson, F.A., Sathe, G., Kane, J.F., 1995. Effects of a minor isoleucyl tRNA on heterologous protein translation in *Escherichia coli*. *J. Bacteriol.* 177, 7086–7091.
- Docimo, T., Reichelt, M., Schneider, B., Kai, M., Kunert, G., Gershenzon, J., D'Auria, J.C., 2012. The first step in the biosynthesis of cocaine in *Erythroxylum coca*: the characterization of arginine and ornithine decarboxylases. *Plant Mol. Biol.* 78, 599–615. <https://doi.org/10.1007/s11103-012-9886-1>.
- Fahey, J.W., Zalcman, A.T., Talalay, P., 2001. The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* 56, 5–51. [https://doi.org/10.1016/S0031-9422\(00\)00316-2](https://doi.org/10.1016/S0031-9422(00)00316-2).
- Galloway, C.A., Sowden, M.P., Smith, H.C., 2003. Increasing the yield of soluble recombinant protein expressed in *E. coli* by induction during late log phase. *Biotechniques* 34, 524–546.
- Garnier, S., 2018. *Viridis: Default Color Maps from 'matplotlib'*.

- George, K.W., Chen, A., Jain, A., Batth, T.S., Baidoo, E.E.K., Wang, G., Adams, P.D., Petzold, C.J., Keasling, J.D., Lee, T.S., 2014. Correlation analysis of targeted proteins and metabolites to assess and engineer microbial isopentenol production. *Biotechnol. Bioeng.* 111, 1648–1658. <https://doi.org/10.1002/bit.25226>.
- George, K.W., Thompson, M.G., Kang, A., Baidoo, E., Wang, G., Chan, L.J.G., Adams, P.D., Petzold, C.J., Keasling, J.D., Lee, T.S., 2015. Metabolic engineering for the high-yield production of isoprenoid-based C<sub>5</sub> alcohols in *E. coli*. *Sci. Rep.* 5, 11128. <https://doi.org/10.1038/srep11128>.
- Geu-Flores, F., Møldrup, M.E., Böttcher, C., Olsen, C.E., Scheel, D., Halkier, B.A., 2011. Cytosolic  $\gamma$ -glutamyl peptidases process glutathione conjugates in the biosynthesis of glucosinolates and camalexin in arabidopsis. *Plant Cell* 23, 2456–2469. <https://doi.org/10.1105/tpc.111.083998>.
- Geu-Flores, F., Nielsen, M.T., Nafisi, M., Møldrup, M.E., Olsen, C.E., Motawia, M.S., Halkier, B.A., 2009a. Glucosinolate engineering identifies a gamma-glutamyl peptidase. *Nat. Chem. Biol.* 5, 575–577. <https://doi.org/10.1038/nchembio.185>.
- Geu-Flores, F., Olsen, C.E., Halkier, B.A., 2009b. Towards engineering glucosinolates into non-cruciferous plants. *Planta* 229, 261–270. <https://doi.org/10.1007/s00425-008-0825-y>.
- Goder, V., Spiess, M., 2003. Molecular mechanism of signal sequence orientation in the endoplasmic reticulum. *EMBO J.* 22, 3645–3653. <https://doi.org/10.1093/emboj/cdg361>.
- Grubb, C.D., Zipp, B.J., Ludwig-Müller, J., Masuno, M.N., Molinski, T.F., Abel, S., 2004. Arabidopsis glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis. *Plant J.* 40, 893–908. <https://doi.org/10.1111/j.1365-313X.2004.02261.x>.
- Guédón, E., Martin-Verstraete, I., 2007. Cysteine metabolism and its regulation in bacteria. In: Wendisch, V.F. (Ed.), *Amino Acid Biosynthesis ~ Pathways, Regulation and Metabolic Engineering*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 195–218. [https://doi.org/10.1007/7171\\_2006\\_060](https://doi.org/10.1007/7171_2006_060).
- Hansen, C.H., Du, L., Naur, P., Olsen, C.E., Axelsen, K.B., Hick, A.J., Pickett, J.A., Halkier, B.A., 2001. CYP83B1 is the oxime-metabolizing enzyme in the glucosinolate pathway in arabidopsis. *J. Biol. Chem.* 276, 24790–24796. <https://doi.org/10.1074/jbc.M102637200>.
- Huang, B., Guo, J., Yi, B., Yu, X., Sun, L., Chen, W., 2008. Heterologous production of secondary metabolites as pharmaceuticals in *Saccharomyces cerevisiae*. *Biotechnol. Lett.* 30, 1121–1137. <https://doi.org/10.1007/s10529-008-9663-z>.
- Huyen, D.T., Giang, L.Q., Hai, T.N., 2015. Expression of flagellin FLJb derived from *Salmonella enterica* serovar typhimurium in *Escherichia coli* BL21. *V. J. Biol.* 36. <https://doi.org/10.15625/0866-7160/v36n4.6180>.
- Ichinose, H., Wariishi, H., 2013. High-level heterologous expression of fungal cytochrome P450s in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 438, 289–294. <https://doi.org/10.1016/j.bbrc.2013.07.057>.
- Ikram, N.K.B.K., Zhan, X., Pan, X.-W., King, B.C., Simonsen, H.T., 2015. Stable heterologous expression of biologically active terpenoids in green plant cells. *Front. Plant Sci.* 6, 129. <https://doi.org/10.3389/fpls.2015.00129>.
- Jeschke, V., Burrow, M., 2018. In: *Glucosinolates*. John Wiley & Sons Ltd, Chichester, UK, pp. 1–8. <https://doi.org/10.1002/9780470015902.a0027968>. eLS. John Wiley & Sons, Ltd.
- Jones, J.A., Vernacchio, V.R., Lachance, D.M., Lebovich, M., Fu, L., Shirke, A.N., Schultz, V.L., Cress, B., Linhardt, R.J., Koffas, M.A.G., 2015. ePathOptimizer: a combinatorial approach for transcriptional balancing of metabolic pathways. *Sci. Rep.* 5, 11301. <https://doi.org/10.1038/srep11301>.
- Juminaga, D., Baidoo, E.E.K., Redding-Johanson, A.M., Batth, T.S., Burd, H., Mukhopadhyay, A., Petzold, C.J., Keasling, J.D., 2012. Modular engineering of L-tyrosine production in *Escherichia coli*. *Appl. Environ. Microbiol.* 78, 89–98. <https://doi.org/10.1128/AEM.06017-11>.
- Klein, M., Papenbrock, J., 2009. Kinetics and substrate specificities of desulfo-glucosinolate sulfotransferases in *Arabidopsis thaliana*. *Physiol. Plantarum* 135, 140–149. <https://doi.org/10.1111/j.1399-3054.2008.01182.x>.
- Kopriva, S., Koprivova, A., 2004. Plant adenosine 5'-phosphosulphate reductase: the past, the present, and the future. *J. Exp. Bot.* 55, 1775–1783. <https://doi.org/10.1093/jxb/erh185>.
- Kredich, N.M., 2008. Biosynthesis of cysteine. *EcoSal Plus* 3. <https://doi.org/10.1128/ecosalplus.3.6.1.11>.
- Laursen, T., Borch, J., Knudsen, C., Bavishi, K., Torta, F., Martens, H.J., Silvestro, D., Hatzakis, N.S., Wenk, M.R., Dafforn, T.R., Olsen, C.E., Motawia, M.S., Hamberger, B., Möller, B.L., Bassard, J.-E., 2016. Characterization of a dynamic metabolon producing the defense compound dhurrin in sorghum. *Science* 354, 890–893. <https://doi.org/10.1126/science.aag2347>.
- Leyh, T.S., Taylor, J.C., Markham, G.D., 1988. The sulfate activation locus of *Escherichia coli* K12: cloning, genetic, and enzymatic characterization. *J. Biol. Chem.* 263, 2409–2416.
- MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., MacCoss, M.J., 2010. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26, 966–968. <https://doi.org/10.1093/bioinformatics/btq054>.
- Miki, Y., Asano, Y., 2014. Biosynthetic pathway for the cyanide-free production of phenylacetone in *Escherichia coli* by utilizing plant cytochrome P450 79A2 and bacterial aldolase dehydratase. *Appl. Environ. Microbiol.* 80, 6828–6836. <https://doi.org/10.1128/AEM.01623-14>.
- Mikkelsen, M.D., Buron, L.D., Salomonsen, B., Olsen, C.E., Hansen, B.G., Mortensen, U.H., Halkier, B.A., 2012. Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab. Eng.* 14, 104–111. <https://doi.org/10.1016/j.ymben.2012.01.006>.
- Mikkelsen, M.D., Naur, P., Halkier, B.A., 2004. Arabidopsis mutants in the C-S lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis. *Plant J.* 37, 770–777. <https://doi.org/10.1111/j.1365-313X.2004.02002.x>.
- Mirza, N., Crocoll, C., Erik Olsen, C., Ann Halkier, B., 2016. Engineering of methionine chain elongation part of glucoraphanin pathway in *E. coli*. *Metab. Eng.* 35, 31–37. <https://doi.org/10.1016/j.ymben.2015.09.012>.
- Miyada, C.G., Stoltzfus, L., Wilcox, G., 1984. Regulation of the araC gene of *Escherichia coli*: catabolite repression, autoregulation, and effect on araBAD expression. *Proc. Natl. Acad. Sci. U.S.A.* 81, 4120–4124. <https://doi.org/10.1073/pnas.81.13.4120>.
- Mizutani, M., Ohta, D., 1998. Two isoforms of NADPH:cytochrome P450 reductase in *Arabidopsis thaliana*. Gene structure, heterologous expression in insect cells, and differential regulation. *Plant Physiol.* 116, 357–367.
- Møldrup, M.E., Geu-Flores, F., Olsen, C.E., Halkier, B.A., 2011. Modulation of sulfur metabolism enables efficient glucosinolate engineering. *BMC Biotechnol.* 11, 12. <https://doi.org/10.1186/1472-6750-11-12>.
- Naur, P., Petersen, B.L., Mikkelsen, M.D., Bak, S., Rasmussen, H., Olsen, C.E., Halkier, B.A., 2003. CYP83A1 and CYP83B1, two nonredundant cytochrome P450 enzymes metabolizing oximes in the biosynthesis of glucosinolates in arabidopsis. *Plant Physiol.* 133, 63–72. <https://doi.org/10.1104/pp.102.019240>.
- Park, S., Kim, Y.-S., Rupasinghe, S.G., Schuler, M.A., Back, K., 2013. Rice P450 reductases differentially affect P450-mediated metabolism in bacterial expression systems. *Bioproc. Biosyst. Eng.* 36, 325–331. <https://doi.org/10.1007/s00449-012-0787-0>.
- Percy, A.J., Chambers, A.G., Yang, J., Domanski, D., Borchers, C.H., 2012. Comparison of standard- and nano-flow liquid chromatography platforms for MRM-based quantitation of putative plasma biomarker proteins. *Anal. Bioanal. Chem.* 404, 1089–1101. <https://doi.org/10.1007/s00216-012-6010-y>.
- Piotrowski, M., Schemenewitz, A., Lopukhina, A., Müller, A., Janowitz, T., Weiler, E.W., Oecking, C., 2004. Desulfo-glucosinolate sulfotransferases from *Arabidopsis thaliana* catalyze the final step in the biosynthesis of the glucosinolate core structure. *J. Biol. Chem.* 279, 50717–50725. <https://doi.org/10.1074/jbc.M407681200>.
- Pontrelli, S., Chiu, T.-Y., Lan, E.L., Chen, F.Y.-H., Chang, P., Liao, J.C., 2018. *Escherichia coli* as a host for metabolic engineering. *Metab. Eng.* <https://doi.org/10.1016/j.ymben.2018.04.008>.
- Pritchard, M.P., Ossetian, R., Li, D.N., Henderson, C.J., Burchell, B., Wolf, C.R., Friedberg, T., 1997. A general strategy for the expression of recombinant human cytochrome P450s in *Escherichia coli* using bacterial signal peptides: expression of CYP3A4, CYP2A6, and CYP2E1. *Arch. Biochem. Biophys.* 345, 342–354. <https://doi.org/10.1006/abbi.1997.0265>.
- Redding-Johanson, A.M., Batth, T.S., Chan, R., Krupa, R., Szmidt, H.L., Adams, P.D., Keasling, J.D., Lee, T.S., Mukhopadhyay, A., Petzold, C.J., 2011. Targeted proteomics for metabolic pathway optimization: application to terpene production. *Metab. Eng.* 13, 194–203. <https://doi.org/10.1016/j.ymben.2010.12.005>.
- Rossi, E., Motta, S., Mauri, P., Landini, P., 2014. Sulfate assimilation pathway intermediate' phosphoadenosine 59-phosphosulfate acts as a signal molecule affecting production of curli fibres in *Escherichia coli*. *Microbiol. (Read. Engl.)* 160, 1832–1844. <https://doi.org/10.1099/mic.0.079699-0>.
- RStudio Team, 2015. *RStudio. Integrated Development for R*, Boston, MA, USA.
- San-Miguel, T., Pérez-Bermúdez, P., Gavidia, I., 2013. Production of soluble eukaryotic recombinant proteins in *E. coli* is favoured in early log-phase cultures induced at low temperature. *SpringerPlus* 2, 89. <https://doi.org/10.1186/2193-1801-2-89>.
- Satishchandran, C., Hickman, Y.N., Markham, G.D., 1992. Characterization of the phosphorylated enzyme intermediate formed in the adenosine 5'-phosphosulfate kinase reaction. *Biochemistry* 31, 11684–11688. <https://doi.org/10.1021/bi00162a003>.
- Siddiqui, M.S., Thodey, K., Trenchard, I., Smolke, C.D., 2012. Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *FEMS Yeast Res.* 12, 144–170. <https://doi.org/10.1111/j.1567-1364.2011.00774.x>.
- Singh, P., Batth, T.S., Juminaga, D., Dahl, R.H., Keasling, J.D., Adams, P.D., Petzold, C.J., 2012. Application of targeted proteomics to metabolically engineered *Escherichia coli*. *Proteomics* 12, 1289–1299. <https://doi.org/10.1002/pmic.201100482>.
- Studier, F.W., 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* 219, 37–44.
- Sønderby, I.E., Geu-Flores, F., Halkier, B.A., 2010. Biosynthesis of glucosinolates—gene discovery and beyond. *Trends Plant Sci.* 15, 283–290. <https://doi.org/10.1016/j.tplants.2010.02.005>.
- Sørensen, H.P., Mortensen, K.K., 2005. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Factories* 4, 1. <https://doi.org/10.1186/1475-2859-4-1>.
- Tan, G.-Y., Zhu, F., Deng, Z., Liu, T., 2016. In vitro reconstitution guide for targeted synthetic metabolism of chemicals, nutraceuticals and drug precursors. *Synth. Syst. Biotechnol.* 1, 25–33. <https://doi.org/10.1016/j.ysmbio.2016.02.003>.
- Traka, M.H., 2016. Health benefits of glucosinolates. In: *Glucosinolates, Advances in Botanical Research*. Elsevier, pp. 247–279. <https://doi.org/10.1016/bs.abr.2016.06.004>.
- Wagner, S., Baars, L., Ytterberg, A.J., Klussmeier, A., Wagner, C.S., Nord, O., Nygren, P.-A., van Wijk, K.J., de Gier, J.-W., 2007. Consequences of membrane protein overexpression in *Escherichia coli*. *Mol. Cell. Proteomics* 6, 1527–1550. <https://doi.org/10.1074/mcp.M600431-MCP200>.
- Wessel, D., Flügge, U.I., 1984. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* 138, 141–143.

- [https://doi.org/10.1016/0003-2697\(84\)90782-6](https://doi.org/10.1016/0003-2697(84)90782-6).
- Wickham, H., 2007. Reshaping data with the reshape package. *J. Stat. Software* 21. <https://doi.org/10.18637/jss.v021.i12>.
- Wickham, H., 2011. The split-apply-combine strategy for data analysis. *J. Stat. Software* 40. <https://doi.org/10.18637/jss.v040.i01>.
- Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.
- Wittstock, U., Halkier, B.A., 2000. Cytochrome P450 CYP79A2 from *Arabidopsis thaliana* L. Catalyzes the conversion of L-phenylalanine to phenylacetaldoxime in the biosynthesis of benzylglucosinolate. *J. Biol. Chem.* 275, 14659–14666. <https://doi.org/10.1074/jbc.275.19.14659>.
- Wycuff, D.R., Matthews, K.S., 2000. Generation of an AraC-araBAD promoter-regulated T7 expression system. *Anal. Biochem.* 277, 67–73. <https://doi.org/10.1006/abio.1999.4385>.
- Yang, D., Cho, J.S., Choi, K.R., Kim, H.U., Lee, S.Y., 2017. Systems metabolic engineering as an enabling technology in accomplishing sustainable development goals. *Microb. Biotechnol.* 10, 1254–1258. <https://doi.org/10.1111/1751-7915.12766>.
- Yang, H., Liu, F., Li, Y., Yu, B., 2018. Reconstructing biosynthetic pathway of the plant-derived cancer chemopreventive-precursor glucoraphanin in *Escherichia coli*. *ACS Synth. Biol.* 7, 121–131. <https://doi.org/10.1021/acssynbio.7b00256>.